

Class 13

Hayoung A15531571

11/9/2021

Reading the tsv file and changing column names

```
blast <- read.delim("mm-second.x.zebrafish.tsv")
```

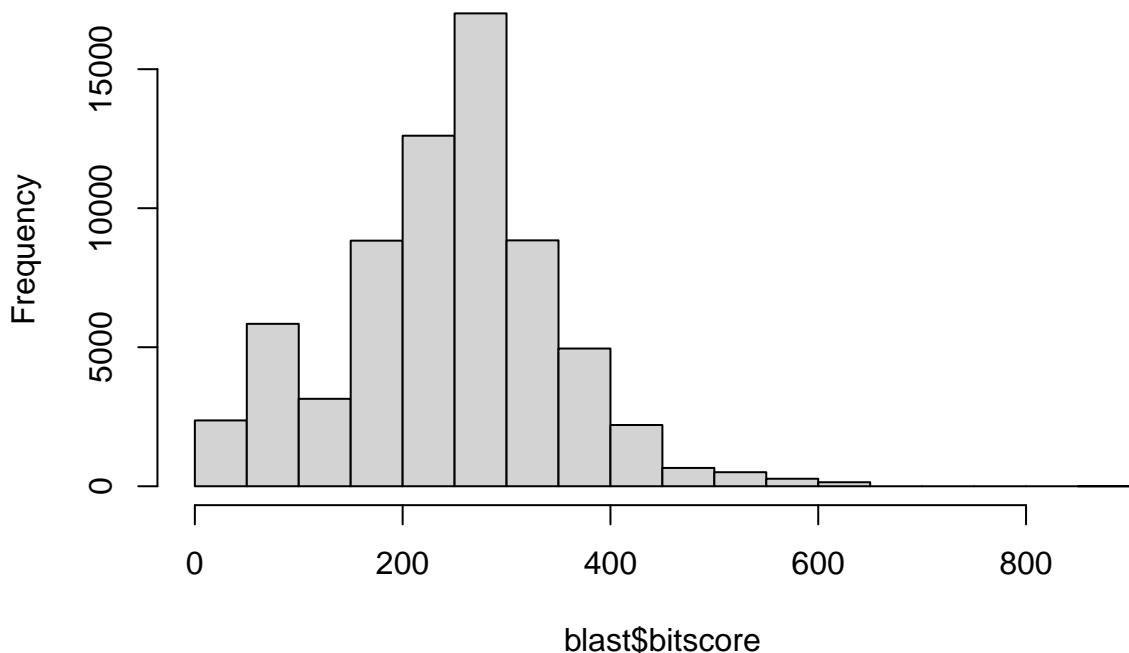
```
colnames(blast) =c("qseqid", "sseqid", "pident", "length", "mismatch", "gapopen", "qstart", "qend", "ss
```

Making a histogram

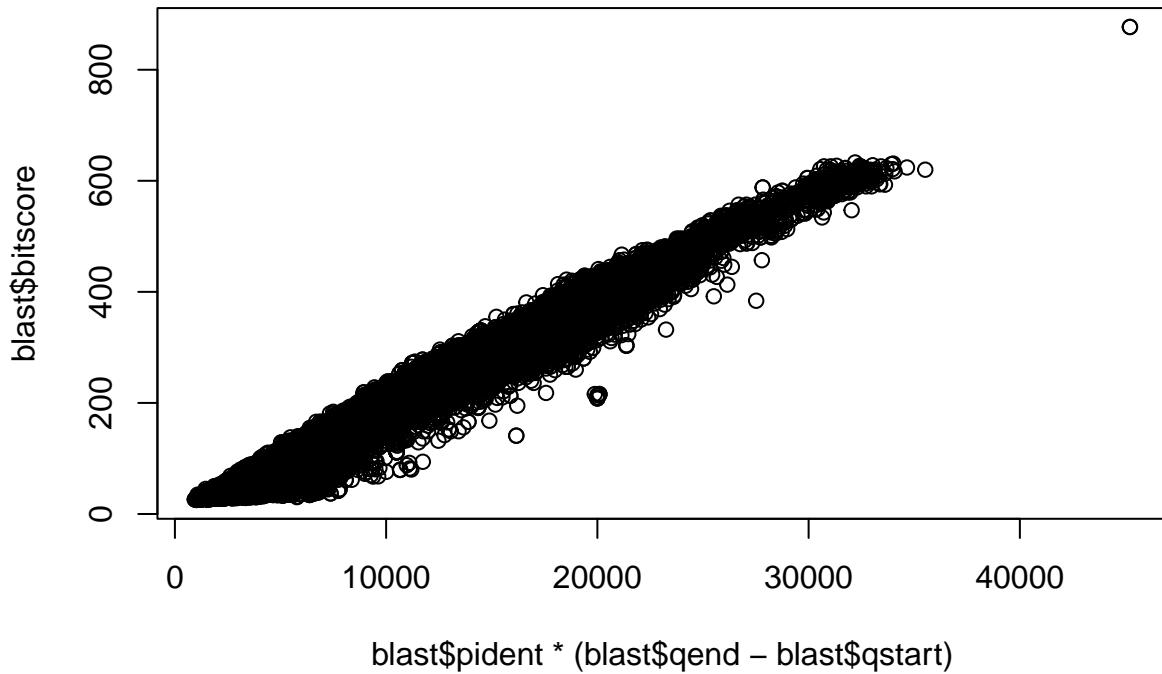
```
library(ggplot2)
```

```
hist(blast$bitscore, breaks=30)
```

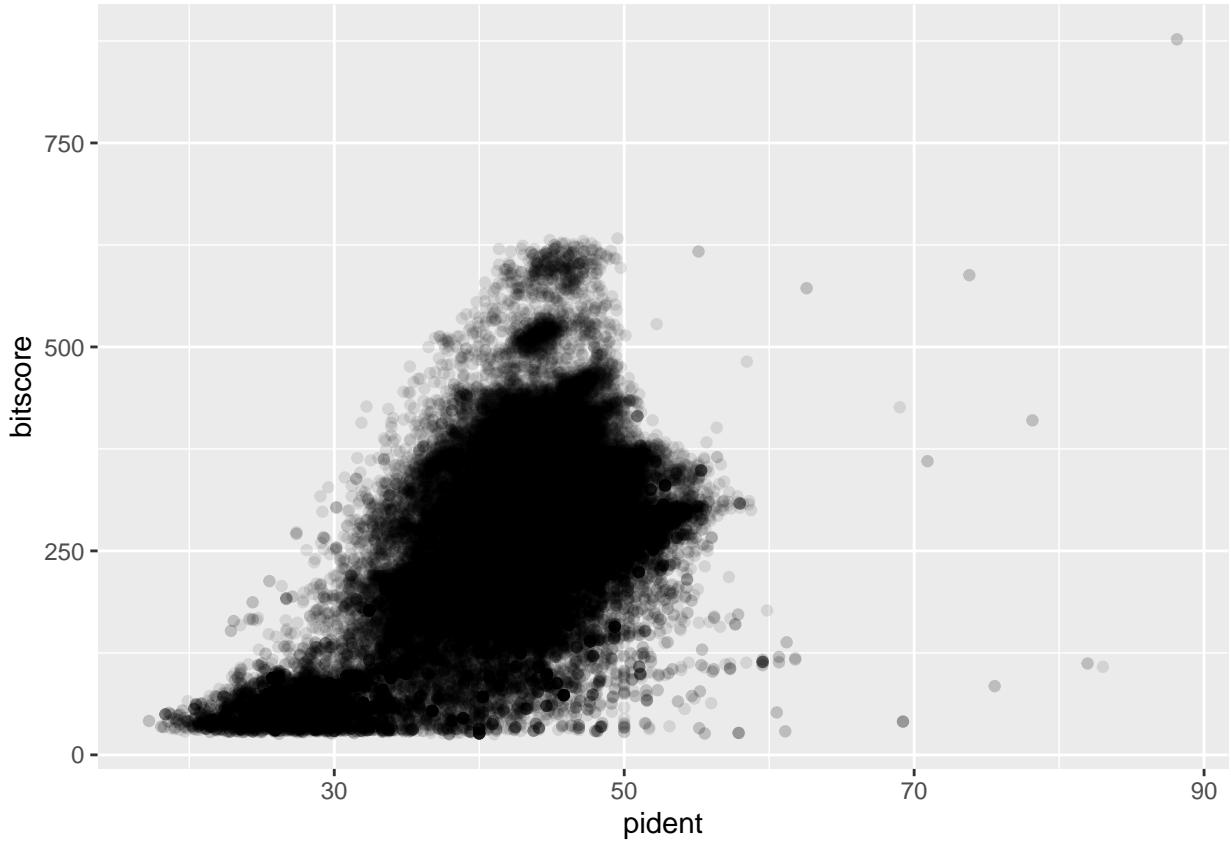
Histogram of blast\$bitscore



```
## Assuming your blast results are stored in an object called 'blast'
plot(blast$pident * (blast$qend - blast$qstart), blast$bitscore)
```



```
ggplot(blast) +
  aes(pident, bitscore) +
  geom_point(alpha=0.1)
```



```
ggplot(blast, aes((blast$pident * (blast$qend - blast$qstart)), bitscore)) + geom_point(alpha=0.1) + geom_smooth(method = "gam", formula = y ~ s(x, bs = "cs"))

## Warning: Use of 'blast$pident' is discouraged. Use 'pident' instead.

## Warning: Use of 'blast$qend' is discouraged. Use 'qend' instead.

## Warning: Use of 'blast$qstart' is discouraged. Use 'qstart' instead.

## Warning: Use of 'blast$pident' is discouraged. Use 'pident' instead.

## Warning: Use of 'blast$qend' is discouraged. Use 'qend' instead.

## Warning: Use of 'blast$qstart' is discouraged. Use 'qstart' instead.

## 'geom_smooth()' using method = 'gam' and formula 'y ~ s(x, bs = "cs")'
```

